DOCUMENT RESUME

ED 075 920 EA 005 049

AUTHOR	Brandes, Barbara J.
TITLE	Problems in the Evaluation of Affective Education: A Case Study.
PUB DATE	Feb 73
NOTE	13p.; Paper presented at American Educational Research Association Annual Meeting (58th, New Orleans, Louisiana, February 26 - March 1, 1973)
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	*Affective Objectives; Case Studies (Education); Educational Objectives; Elementary School Students; *Formative Evaluation; *Goal Orientation; *Instructional Materials; Measurement Techniques; *Research Methodology; Speeches

ABSTRACT

Critical methodological issues in the design and evaluation of programs in affective education are discussed. The uncertain relationship between the short term and long range goals of instruction is problematic to evaluation in the traditional subject matter areas; but it is a much more serious predicament for evaluators of education in values and attitudes where little is known about the skills essential to the attainment of long range goals. Procedures being used for formative evaluation of an elementary school program in achievement behavior are presented. These procedures illustrate ways of handling the above problem by allowing for successive shaping of objectives. (Author)

PROBLEMS IN THE EVALUATION OF AFFECTIVE EDUCATION: A CASE STUDY[1]

Barbara J. Brandes
Research for Better Schools, Inc.

Education is future-oriented. Virtually all instructional endeavors are aimed at longrange, diffuse goals such as competence in basic mathmatics or proficiency in reading comprehension. Since none of these longrange goals can be attained all at once, we break up areas of knowledge into components and sequence them and hope that the cumulative effect of the components will result in the desired goal. But until we have performed a complete hierarchial analysis of the skills which produce the desired end product, we don't know for sure that we have properly sequenced our components or included instruction on all essentials. This uncertain relationship between longrange goals and outcomes of short-term instruction is problematic to evaluation in the traditional subject matter areas, but it is a much more serious predicament for evaluators of education in values and attitudes where we know little about the skills essential to attainment of longrange goals or whether our goals even consist of a discrete set of skills.

This paper presents results of a small-scale formative tryout of the curriculum package, Achievement Competence Training (ACT), which is being developed at Research for Better Schools, Inc. to promote achievement behavior among late elementary school children. The longrange, global objective of the program is to facilitate students' success in attaining their personal goals. The strategy used in design and formative evaluation of the ACT package is offered as a case study illustrating ways of dealing with some of the problems in affective education.

---

1 Paper presented at the annual meeting of the American Educational Research Association, New Orleans, February, 1973.

I. Some Intractable Problems for Affective Education

Most of the problems which we encounter in designing and evaluating affective programs are not really unique to the affective area. However, it seems to occur more often for affective than for other types of programs that empirical data useful for improving programs are not obtained. It is a frequent occurance for participants in a program to be subjectively convinced that the program has had an impact when there are no empirical data to substantiate this feeling. Either the wrong things have been measured in the evaluation or the right things have been measured in the wrong way. In either case the data are not useful for demonstrating that the program worked or for suggesting how the program might be made to work better.

A. The Problem of Making Program Content Consistent with Program Goals.

One can only really teach that which is teachable. A position which I ascribe to is that to be teachable, a behavior must be capable of being analyzed into a specific hierarchy of skills. Because this feat is difficult or impossible for many affective goals, we end up with programs in which there is little relationship between the content of the program and the statement of program goals. Program content should have some explicit or implicit relationship to expected outcomes.

B. The Problem of Goal-Relevant Measures of Affective Outcomes.

In affective education we are not concerned just with whether students are able to do certain things but more with whether they actually do those things by choice. This fact makes for the greatest challenge in affective education because it requires both that the treatment be powerful enough to alter behavior patterns and that the measurements of outcomes be ingenious enough to assess behavior dispositions in a free-choice context.

These free-choice measures are what is meant by goal-relevant measures of affective outcomes.

C. The Problem of Content-Relevant Measures of Affective Outcomes.

Outcome measures in any program ought to reflect the extent to which students have mastered whatever was taught in the program. Mastery testing of this sort is the only valid method for determining areas in which instruction needs to be improved. It sometimes happens in affective education that the outcome measures reflect the stated goals of a program but bear little resemblance to the actual instructional content of the program. Measures of mastery are what is meant by content-relevant measures of affective outcomes.

The problems outlined here might become more salient in the context of a simple example. This particular example is discussed by Popham in one of the AERA monographs on curriculum evaluation. [2] The example concerns a teacher who believed that her course would help students to become better citizens. However, the course consisted only of a set of facts, and student mastery was assessed on a true/false test. In this example the outcome measures are consistent with the content of the course, but neither the instructional content nor the outcome measures are consistent with the stated goals of the course. The teacher obtains no information about whether she really is helping her students to become better citizens.

2. Popham W. James. Objectives and Instruction. In W. James Popham et al., Instructional Objectives. AERA Monograph Series on Curriculum Evaluation. No. 3. Chicago; Rand McNally, 1969

Popham recommends that the teacher restate her goals to make them consistent with the content of the course -- i.e., the objective is for students to list essential facts about former Presidents, or some such. In fact, a proper evaluation of the course should encompass not only students' learning of the actual content of the course but also the relationship between mastery of the content and the intended outcome; to help students to become better citizens. If the teacher loses sight of her true goal and becomes totally absorbed in how to better teach facts about past Presidents, she may cease looking for other and better methods of helping students to become better citizens. If the desired outcomes of a program actually exist on several levels, they should be stated as such, and appropriate measures for each level should be devised.

II. Dealing with Some of the Problems in Affective Education

If, as is the case in most areas of affective education, there is too little known about the behaviors in question to make them truly teachable, then we must acknowledge that we have a complex situation in which we must undertake several things simultaneously. The complexity of the situation should be reflected both in the statements of goals and in the measures used to assess outcomes. The strategy advocated here is to formulate short-term lesson and program objectives, to obtain measures of their attainment, and to attempt to relate short-term outcomes to increments of progress toward long range goals.

A. The Procedure of Deriving Program Goals from Longrange Goals.

Statements of program goals must be made consistent with reasonable expectations given the length of the program and whatever is known about the plasticity of the behaviors in question. Usually this means that a program must focus on a few fairly specific behaviors which are believed to have some bearing on the affective outcomes desired in the long term. In other words, it is necessary to perform a task analysis of the longrange affective goals. To continue with Popham's example of the teacher who wanted to help her students to become better citizens, there is clearly a need for a reductionist process that will lead to a statement of program goals which is both specific and appropriate to the estimated power of the intervention program being planned. The teacher obviously did this implicitly because she derived both a course of instruction and procedures for assessing outcomes. However, the derivation should be made explicit and optimally should reflect the most sophisticated judgment possible given the present state of knowledge about the behaviors in question.

B. The Procedure of Deriving a Hierarchy of Outcome Statements.

Beyond deriving a statement of program objectives from longrange goals, it is also necessary to formulate a clear hierarchy of hypotheses about outcomes which should follow from the instructional program. This procedure provides a set of testable and hierarchically ordered hypotheses. At the lowest level is an hypotheses about the outcomes which should result from a given segment of the program. At the next higher level is a statement about the set of behaviors which should indicate mastery of the entire instructional program. Next should follow a statement about ways in which students completing the program would be expected to behave differently from other students. At this level an assertion is made about the relationship of the instruction provided

to the behavior changes being sought in the program. Finally there should be a statement asserting the relationship of these behavioral outcomes to the longrange goals from which the program was derived.

C. The Procedure of Devising Assessment Techniques Appropriate to Each Level of Outcomes.

In Table 1 the four levels of outcomes and the assessment techniques appropriate to each level are presented. The first two levels are those which are most frequently discussed in prescriptions for evaluation. Students should

TABLE 1

Assessing Outcomes of Instruction at Multiple Levels

Level	Outcome	Target of Assessment	Method of Assessment
1	Students master lesson or unit content	Lesson or unit instructional content.	Criterion-referenced lesson or unit mastery test.
2	Students master program content.	Instructional content and sequence for entire program.	Criterion-referenced program mastery test.
3	Students display specified behaviors more often than their agemates.	Relationship of instructional content to desired behaviors.	Norm-referenced test with comparison group.
4	Students display behaviors consistent with longrange goal at adulthood.	Relationship of behaviors displayed by students at end of program to longrange goals.	Longitudinal study or informed judgment based on synthesis of available data.

be tested following units of instruction and at the end of the entire program to determine their mastery of material presented to them for learning. Test items tend to be direct extrapolation of the content of the instructional program.

Students' actual performance can be compared with an a priori established level of acceptable performance, and a decision can be made about whether to revise portions of the program. In Popham's example the true/false test administered at the end of the program was an attempt at this type of mastery testing.

At level 3 an attempt is made to verify the relationship between the instruction provided in the program and the behaviors which the program is trying to promote. In Popham's example it was the teacher's neglect of this level which was the weakest part of her program. She should have attempted to find out whether her students displayed some of the behaviors of good citizenship more often than other students. Adequate assessment at this level requires comparison-group testing which is not biased in terms of vocabulary or peculiar content of the instructional program and which hopefully gets at behavior dispositions rather than just content mastery.

At the fourth and final level the relationship between behavioral outcomes of the program and the longrange goal is reassessed. Since longitudinal evaluation studies are typically not feasible and unlikely to reveal effects of a single program, one must usually rely on his own informed judgment of the value of program outcomes. However, a thorough evaluation should include a reassessment of program goals based on an updated synthesis of research data about the behaviors in question.

III. Case Study of an Affective Program: Achievement Competence Training

The instructional package Achievement Competence Training (ACT)[3] is being developed as part of the Humanizing Learning Program at Research for Better Schools, Inc., to instruct fifth grade children in the skills and attitudes of goal attainment behavior. The long-range educational goal of the program is to facilitate students' success in attaining their personal goals.

The first step in developing an instructional product was to derive specific program goals or focal points from this global long-range goal. The procedure was to look at research information about the criterion population (successful achievers) and to attempt to select teachable skills from their behavioral repertoire. An attempt was made to describe the distinguishing dispositional qualities of high achievers and to generate a behavioral sequence which describes the process of goal attainment. This effort culminated in a six step strategy of goal attainment which became the core of the ACT program. Thus the synthesis of information about the criterion population resulted in two general program goals: (1) to teach and encourage the use of the ACT strategy of goal attainment, and (2) to enhance positive self-evaluation and belief in internal control. Objectives for mastery of program and component (lesson, unit) content were arrived at by conceptual analysis of the goal attainment strategy so as to provide a thorough exposition of each step and repeated practice in the use of the strategy.

---

3. The Achievement Competence Training materials are being developed by Dr. Russell Hill and his associates in the Humanizing Learning Program, Research for Better Schools, Inc.

A. Stating Objectives and Measuring Outcomes at Multiple Levels: Medium Risk Taking.

One indication of proficiency in the use of the ACT achievement strategy is the level of aspiration or risk reflected in the goals set by students. Students successfully completing the ACT program should set goals for themselves which are challenging for them but not so difficult as to be impossible to attain. The term for this quality of personal goals is "medium risk." Instruction in medium risk-taking was incorporated in the ACT program because of research evidence indicating that high achievers are able to integrate knowledge of their own past performance with their desire for excellence and thereby set goals at a level of difficulty consistent with, or slightly higher than, their past performance. Low achievers, on the other hand, often set goals which are too easy or impossibly difficult.

The process of deriving a hierarchy of objectives pertaining to medium risk taking from the long-range goal of ACT results in the levels of outcomes depicted in Table 2. Mastery objectives, or the teachable content associated with medium risk-taking, include providing a definition of medium risk, discriminating medium risk goals from low and high risk goals, recognizing medium risk as a quality of goals set by achievers, and so forth. The right-hand column in Table 2 shows the means by which attainment of these objectives is assessed. At the lesson and program mastery levels (levels 1 and 2 in Table 2) assessment procedures directly reflect the objectives as stated and are quite amenable to criterion-referenced testing procedures. At these levels of assessment, we are concerned with whether or not the instruction provided in the ACT program is adequate to teach students information about medium risk-taking. If students fail to grasp and retain this information, the program developers need to revise the sections on medium risk-taking in order to provide clearer explications, more practice or remedial materials.

TABLE 2

Assessing Levels of Outcomes for One ACT Goal: Medium Risk-Taking

Level	Outcome	Target of Assessment	Method of Assessment
1	Students master concept of medium risk-taking.	Instructional content of lesson on medium risk-taking.	Lesson posttest.
2	Students retain information about medium risk-taking.	Instructional content pertaining to medium risk-taking for whole program.	Program mastery test.
3	Students take medium risks more often than their agemates.	Relationship of instructional content on medium risk-taking to frequency of taking medium risks.	Situational tests of risk-taking comparing ACT with non ACT students.
4	Students have more success than their agemates in attaining personal goals at adulthood.	Relationship of medium risk-taking to success in attaining personal goals.	Updated synthesis of research data on achievement behaviors.

At level 3, the concern is with the effects of instruction on unprompted behavior. Once students have learned about medium risk goal-setting, are they actually more likely to set medium risk goals for themselves when compared with their agemates who have not been similarly instructed? Assessment at this level requires selection of one or more indicators of medium risk goal-setting. In the ACT evaluation, ACT and non ACT students are compared on two situational tests of risk-taking.

Finally at level 4, we again ask whether or not teaching students to set medium risk goals does actually further their success in attaining personal goals. We cannot obtain direct and unequivocal evidence pertinent to this question, but we can look at several types of evidence to get an

idea of whether we are on the right track. First, we can look at how ACT students perform on other indices of achievement. If we find the desired effects on some of these other indices, we can be reassured that instruction in medium risk goal-setting has some coherence with other achievement behaviors. Second, we can reexamine the growing body of basic research data on achievement behavior to see whether medium risk-taking, as taught in ACT, still seems to be a significant component of the achiever's repertoire.

B. Obtaining Evidence of Coherence Between Levels of Outcomes: Perceived Locus of Control.

It is typically the case during formative evaluation of a program that significant effects on affective outcomes are not attained. It is also generally true during early stages that student performance on the measures of content mastery (levels 1 and 2 in Table 1) fall short of the criterion for acceptable performance. However, information from these early formative tryouts is most useful if it can provide some hints of a relationship between content mastery and affective outcomes even though it is premature to obtain a direct test of the efficacy of the instructional content in fostering the desired affective outcomes. Some relevant presumptive evidence can be obtained by looking at correlations between student performance at the various levels of outcomes. In the formative evaluation of ACT we have attempted to gain such evidence pertaining to the goal of enhancing internal locus of control.

During the 1971-72 school year a small-scale tryout of the ACT program was conducted in a suburban Philadelphia school district. Two classes of fifth grade students received the entire sequence of 24 lessons, and two other matched classes served as an uninstructed comparison group. Among the results favorable to the existing instructional materials and procedures was a significant effect showing higher performance of experimental-group students on a test designed to measure some of the cognitions involved in self-directed behavior. On the other hand, the effect did not attain significance on the measure of belief in internal control (Bialer Children's Locus of Control Scale). Although we must conclude at this point that there is no evidence that the ACT program enhances belief in internal control, we did note that performance on the test of cognitions involved in self-directed behavior correlated significantly with posttest internal control. In a multiple regression analysis with posttest internal control as the dependent variable, the cognitive measure was found to account for the largest portion of variance after pretest internal control was entered. (Multiple R for these three variable was .71). Neither IQ nor academic achievement were significant predictors of internal control. This result suggests to us that what is taught in the program, as reflected on the measures of cognitions related to self-direction, may have some validity with respect to internal control.

Some additional corroborative evidence comes fron an analysis within the sample of children receiving the ACT program. Again using posttest internal control as the dependent variable in a multiple regression analysis, the best predictors in descending order were pretest internal control, the measure of cognitions pertaining to self-direction, and the ACT Mastery Test (end-of-program test assessing mastery of concepts taught in ACT). (Multiple R for all four variables was .70). Students who mastered program content best were, to some extent, the same students who did well on the cognitive measure and on posttest

internal control. On the basis of these data we have concluded that some components of self-directed behavior were successfully taught by the ACT program. Of course the final test of the effect of the program on perceived locus of control must be made directly.

Results of the pilot tryout were used to make refinements both in the instructional content of ACT and in the procedures used to assess outcomes. The revised ACT program is undergoing a large-scale field test and evaluation in 33 Philadelphia area schools during the 1972-73 school year.